In: Proceedings of the 18th International Association of Forensic Phonetics and Acoustics Annual Conference, August 2009, Department of Linguistics, University of Cambridge, U.K.

Ratios of spectral energy bands can indicate speech volume level

Anja Geumann
Department of Phonetics, Johann Wolfgang Goethe-Unversity, Frankfurt, Germany
geumann@em.uni-frankfurt.de

Spectral patterns in loud speech are compared to speech produced at normal volume. It will be argued that these could more robustly indicate the actual volume in speech than volume recorded, which in turn most often is insufficient to reconstruct volume at source. On the other hand, volume at source is known to have an effect on characteristics important for speaker identification, such as fundamental frequency, sound duration, and first formant frequency (see e.g. Jessen/Köster/Gfroerer, 2005, Traunmüller/Eriksson, 2000).
In our experiment, data from consonants in different vowel contexts were analysed. All phrases were produced in loud and in normal speech. Analyses were performed with the Praat analysis program for the following parameters: Energy in Pa^2/s of different spectral bands between 0.5 and 10 kHz, as well as total energy, in addition to normalised band energy (E_{ratio} explained below). Spectrograms were obtained by FFT as far up as 10 kHz. Energy in each spectral band was determined from a spectral slice of app. 30 ms length taken in the middle of the sound. In a first experiment, for the alveolar and post-alveolar fricative, normalised, or relative, energy was calculated as described in (1) and (2), where E_{total} is Energy in a band ranging from 0.5-10 kHz.

(1) $E_{rel} = E_{band} / E_{total}$

(2) $E_{ratio} = \text{loud } E_{rel} / \text{norm } E_{rel}$

(3) $\text{VtoC\_}E_{ratio} = \text{V\_}E_{rel} / \text{C\_}E_{rel}$

The overall impression from band energy in loud speech is, of course, that overall energy is much higher. However, additionally some spectral bands are raised disproportionally.
The overall finding is that spectral distributions captured with parameters such as centre of gravity for fricatives remain fairly constant in loud speech. What is more, some energy bands actually decrease in proportion, in loud speech.
In a second analysis, the ratio between vowel /a/ and the following alveolar or post-alveolar fricative is analysed for the same energy bands as mentioned above. The VtoC energy ratio in loud speech is disproportionally increased between 0.5-1.5 kHz, i.e. the spectral range covering the first two formants of the vowel. It will be argued that the ratio between energy of different spectral bands and between (certain) adjacent sounds can be interpreted as indicator of volume level and can be an useful help in estimating volume level to correct parameters for speaker identification as f_0 (Jessen/Köster/Gfroerer, 2005) or formants (Nolan/Grigoras, 2005).

References

Boersma, P. and D Weenink. (2007). Praat: doing Phonetics by Computer (Version 4.5.16) [Computer Program] Retrieved February 18, 2007 from http://www.praat.org.

Jessen, M., O. Köster and S. Gfroerer (2005). Influence of vocal effort on average and variability of fundamental frequency. International Journal of Speech, Language and the Law, vol. 12, 174-213.

Nolan, F and C. Grigoras (2005). A case for formant analysis in forensic speaker identification. International Journal of Speech, Language and the Law, vol 12, 143-173.

Traunmueller, H. and A. Eriksson (2000). Acoustic effects of variation in vocal effort by men, women, and children. J. Acoust Soc Am, 107, 3438-3451.